LA-UR-00-1977

Statistics for Complex Computer Models: Beyond Input-Output Analysis

Katherine Campbell and Richard Beckman

Statistical Sciences Group

Los Alamos National Laboratory

http://www.lanl.gov/orgs/tsa/tsa1/index.html

Presented at Interface 2000, New Orleans, April 6, 2000

This talk was presented as part of an invited session on Critical Infrastructure Models. The emphasis of the other talks was on the theory and application of simulation models. At first glance, the simulation models described by the other speakers in this session might appear to be radical departures from more familiar types of numerical models of dynamical systems, such as models of oil reservoirs or accelerators, based on solving a set of partial differential equations. A closer look at the structure of such models, however, reveals more similarities than differences. Exploring these common features, we find a great variety of statistical opportunities, from the stochastic simulation of incompletely known input and boundary conditions to the design of experiments to optimize computational algorithms, from model calibration to model assessment.

Acknowledgements

- The TRANSIMS team, especially the “feedback” team including Brian Bush, Stephen Eubank and Jim Smith
- LANL ocean modelers including Bob Malone and Rick Smith
- Andy Wolfsberg (Cl-36 work)

I’m indebted to many people for slides and ideas that I’ve incorporated into this talk, including the entire TRANSIMS team but more especially a subteam that is looking at the question of feedback within TRANSIMS; ocean modelers at Los Alamos who are generous with their time and graphics; and my colleague Andy Wolfsberg with whom I have collaborated in some modeling and data analysis for Yucca Mountain.

Outline

I. Introduction: TRANSIMS vs. a “traditional” computer model

II. Statistical simulation of model inputs

III. Model calibration

IV. Model assessment

V. Conclusions

After an introduction, this talk will focus on three areas in which there is a great deal of scope for statistical work in the enterprise of creating and using computer models.

I. Complex computer models

- Dynamical
  - Some kind of process model
    - "Traditional": Systems of differential equations
    - "Novel": Cellular automata, sequential dynamical systems...
- Composed
  - Maybe on more than one scale
    - Very large ensembles of similar, locally interacting components
    - Coupling of several dissimilar components
- Large
  - Large fields of input parameters
  - Huge amounts of output
  - Ensemble dynamics less well understood

At the heart of a complex computer model of the type we are talking about is some computer representation of a dynamical process. As statisticians, we are perhaps not too much involved in the details of this core model. And in some essential respects the details of how the dynamics of the system are abstracted for computational purposes does not affect the basic issues that surround the subject of computer modeling.

Typically, for a computer model that we would consider "complex", the model may be "composite" on more than one scale. TRANSIMS and other infrastructure models tend to use large ensembles of similar, relatively simple components, among which the interactions are localized in space and/or time. Earth modeling is moving in the direction of coupling together large submodels interacting across extensive interfaces (which themselves often need to be constructed as explicit submodels.)

Perhaps the most interesting aspect of "large" is the last bullet. Whether we put together large ensembles of simple components or couple together a few complex models, unexpected things happen. In the last few years, for example, we have seen the suggestion that global warming could switch off the thermohaline circulation in the North Atlantic rather rapidly emerge from a preliminary coupled ocean-atmosphere model.

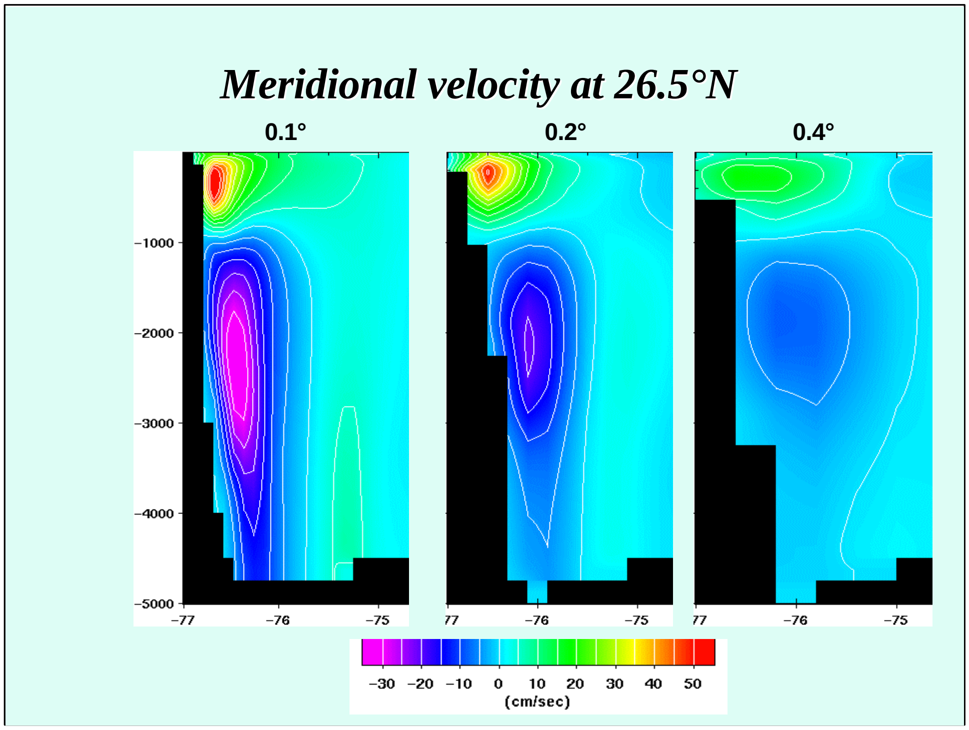
Our intuition may fail us when it comes to predicting what will happen. On the other hand, the results very often help us develop new and useful intuition.

Requirements for ocean model

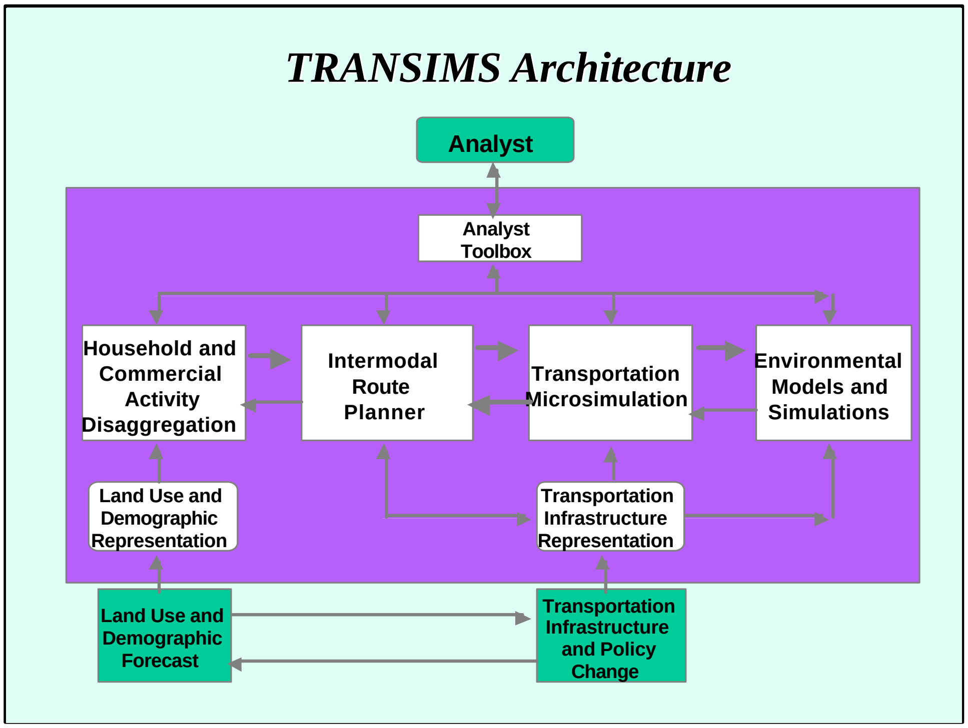
- Representation of dynamics: PDE system
- Boundary conditions
  - Basin topography (static)
  - Wind fields (dynamic)
- Calibration
  - Parameters of subgrid phenomena
  - Spin-up to initial conditions
- Assessment
  - Comparisons with satellite data (SSH for example)
  - Inspection

This outline describes complex computer models in terms that are perhaps more relevant to a statistician rather than a computer modeler. Let me go through this outline using a model of the “traditional” kind, namely an ocean model. First, as already mentioned, there is some kind of representation of the dynamics of the system; for an ocean model this consists of a fairly well-established systems of PDEs. But of course these PDEs have to be reduced to numerical form; this means choosing a grid and so forth, and most importantly, it almost always means making some critical decisions about how to parameterize subgrid (unresolved) phenomena (eddies, in ocean models.) So already things are not so simple, but this is only the beginning.

All computer models have boundary conditions, and usually both static and dynamic. Generally, to say that these boundary conditions, which are entire <u>fields</u> of parameters, are poorly known is a gross understatement, although even if well known they may be poorly represented in the model. Then we come to the interesting topic of model calibration. Here this term refers not only to the parameters of those subgrid physics models but also very often to refining those uncertain boundary conditions and also, in the case of ocean models, to generating an acceptable initial state for the model, one that is compatible with the basic dynamics. And then, finally, when our model is built and calibrated, we need to demonstrate somehow that it is usable for whatever purpose we have in mind; to build confidence in it by one or preferably several “validation” exercises. This is perhaps hardest of all.



This is an example of what the ocean modelers worry about. Model resolution turns out to be an important issue. They tell me that until you get down to about 0.1 degree, models are "wrong by inspection"; no elegant statistics are needed to demonstrate the problems. This series, which really should be read from right to left, shows the development of the deep "return current" (toward us, out of the page) beneath the North Atlantic below the Gulf Stream (the strong red current into the page.) Notice how even the 0.1 degree model, which is pretty much state of the art and requires the computational power available at Los Alamos, still has a very crude approximation to basin topography. Nonetheless topography can be very important; a reasonable representation of the southern Caribbean is essential to producing currents that flow in the right direction farther north, near Cuba.



In the context of TRANSIMS, this outline is realized as follows:

The dynamical core model is the Transportation Microsimulator, a representation of the dynamics of traffic.

This model requires a representation of the transportation infrastructure for the region, which means not only the physical systems of roads and rails but also schedules and other amenities and incentives that may be built into the system. These are the static boundary conditions.

Feeding into the model are representations of demand for transportation services based on the particular region being modeled (the dynamic boundary conditions), which are translated into plans to be executed by the Microsimulator by a separate piece of code, the Router.

Output can consist not only of detailed traffic simulations and summaries but also, an important part of this architecture from the point of view of our customer, detailed simulations of emissions and environmental effects. We must search among such immediate and post-processed outputs for results that can be used for model assessment.

Modeling with TRANSIMS

- Representation of dynamics
  - Small-scale dynamics (microsimulation)
  - Ensemble dynamics (emergent)
- Boundary conditions
- Calibration
- Assessment

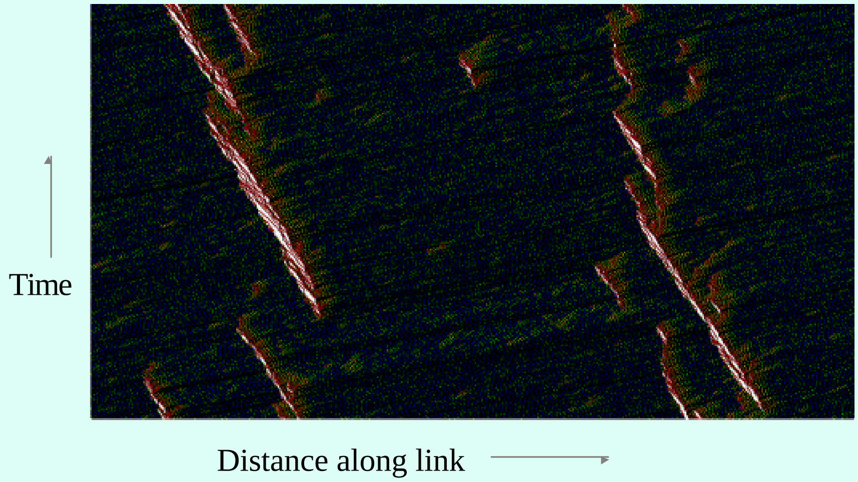
I won't spend much time on the dynamical aspects of TRANSIMS. But it's useful to distinguish dynamics on two or even three levels. The small-scale dynamics of traffic are explicitly embodied in the microsimulator. Medium-scale dynamics--traffic patterns--emerge when this simulator is run. Still larger-scale dynamics--the patterns of interaction between the travelers and the infrastructure--also emerge if the system is allowed to evolve with respect to some of the boundary conditions. The utility of TRANSIMS results from this evolutionary potential, and it is the quality of such evolved patterns that require assessment.

Cellular Automaton Driving Rules

- total of about twelve adjustable parameters for driving rules
- movement forward on grid based on gap to next vehicle, current speed, maximum speed
- lane changes based on chosen approach lane to next intersection, current speed, gap to next vehicle in current lane, gaps to previous and next vehicles in new lane
- intersection entry based on position and speed on link, occupancy of intersection buffer, state of oncoming and interfering traffic

The dynamical core is a cellular automaton that abstracts the basic rules of traffic on networks. There are deliberately a minimal number of parameters to be determined in this part of the system, quite comparable to the number of parameters describing the basic dynamics of the ocean model.

Example Vehicle Trajectories



This is an example of “emergent dynamics” on the middle scale, at the level of traffic patterns. What’s actually plotted is the position of individual vehicles along a one-dimensional traffic link as a function of time, but if you defocus a bit (if you have a choice) and think of this as a pseudo-color map of traffic density instead, this illustration is readily interpretable as showing how random traffic slowdowns on a freeway propagate upstream.

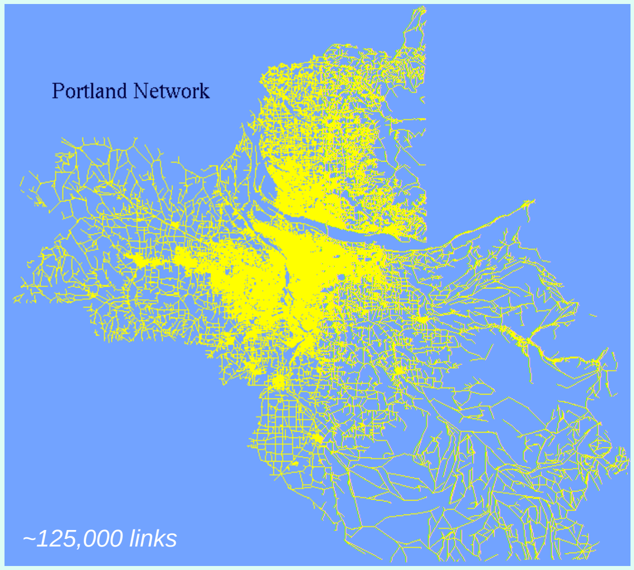
II. Modeling with TRANSIMS: Boundary conditions

- Representation of dynamics
- Boundary conditions
  - Static conditions
    - Network, transit routes and schedules
  - Dynamic conditions
    - Realizations of population, demand
- Calibration
- Assessment

The first topic where I think statisticians really have a lot to contribute is in completing the specification of boundary conditions. Again, here we are talking about fields of parameters, heterogeneous and typically sparsely observed by comparison with the scale of the model, which must nevertheless be completely specified before we can run the model. This specification requires statistical modeling and inference.

(Blue indicates a topic I'm going to talk about in some detail.)

Example Network for Portland, Oregon



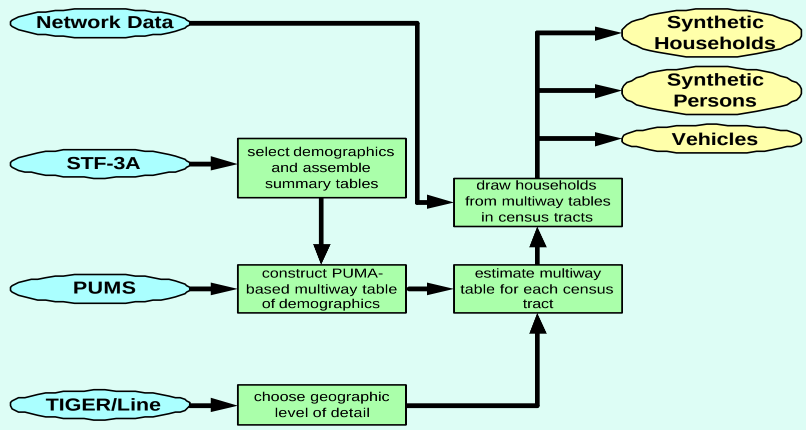
But first, to reiterate, the static boundary conditions for a TRANSIMS model consists of the road and rail networks, the bus routes and schedules, and also things like parking and pedestrian policies in the downtown area. These can be summarized as “the level of service” provided by an existing or projected transportation system. In theory these are probably well known (like basin topography for an ocean model). In practice, being sure that they are correctly or at least adequately represented is labor-intensive work. There are also problems associated with generating future networks for future scenarios which will probably have a statistical component.

Population Synthesizer: Purpose

- creates a regional population realization...
  - demographics closely match real population
  - households are distributed spatially to approximate regional population distribution
- synthetic population's demographics form basis for individual and household activities requiring travel
- household locations determine some of the travel origins and destinations

Statistical modeling is even more important when a time dimension is added. The dynamic boundary conditions for a regional transportation model constructed using TRANSIMS is the pattern of demand for the transportation system. The first step in simulating this demand is to synthesize a population based on census data.

Population Synthesizer: Algorithm (Beckman, Baggerley and McKay 1996)



The census data are marginal data by census block group, supplemented by Public Use Microdata Samples for areas that includes many block groups. What the model needs is a complete, spatially distributed population by household, down to the individual level. Households must be located relative to the network (that is, the static boundary conditions) and assigned personal vehicles. Individuals within households must be assigned characteristics that will be important in determining the types of demands they make on the system, in a way that is consistent with the known or projected marginal demographics.

Population synthesizer: IPF steps

Contingency table	Margins	RAKE'd estimate
Weighted cross-classification of PUMS data by STF-3A marginal variables	Sums of STF-3A margins across census block groups in PUMA	"Average" table for PUMA
Table of ones with one more dimension than number of marginal variables	Individual block group margins plus the PUMA "average" table	Complete table for each block group in the PUMA
Weighted cross-classification of PUMS data by forecast variables	Individual block group margins for the forecast variables	Forecast table for each block group in the PUMA

Base year

Forecast year

The main statistical tool used here is the version of iterative proportional fitting known as the RAKE algorithm. For a base or census year this is a two-step algorithm described by a paper in Transportation Research. The variables used in the TRANSIMS population synthesizer are things like the age and race of the householder, the number of workers in the household, and household income.

Data on these variables from the public use microdata sample are RAKE’d against the census margins or projected margins.

Example Household from PUMS in Portland, Oregon

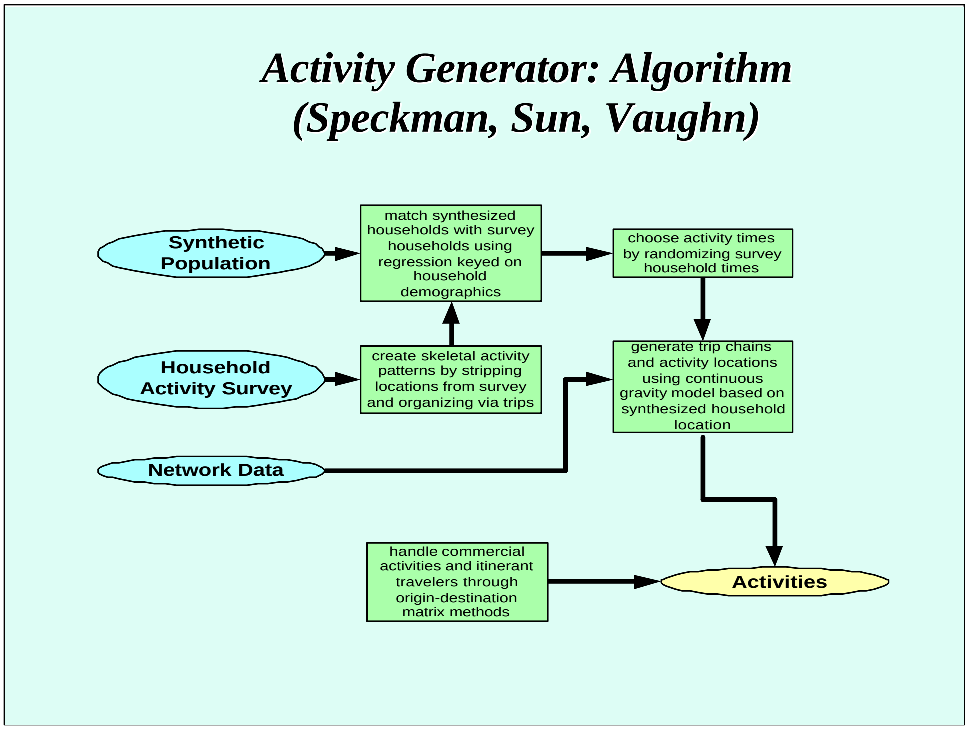
Age	26	26	7
Income	$27k	$16k	$0
Status	worker	worker	student
Automobile			

The result is a table of proportions for resampling the Public Use Microdata Sample within a given census block. The sampled households are located within the census block in a second step. So now we have the potential users of the transportation network.

Activity Generator: Purpose

- creates . . .
  - household and individual activities
  - activity priorities
  - activity locations
  - activity times
  - mode and travel preferences
- generates travel demand sensitive to demographics of synthetic population
- activities form basis for determining individuals' trip plans for the region

The next step is to generate specific activities for each household and individual that result in demands for transportation: travel to work, to school, shopping, visiting, etc. Like the synthesis of population, this step too is based on devising a resampling scheme to make use of available microdata.

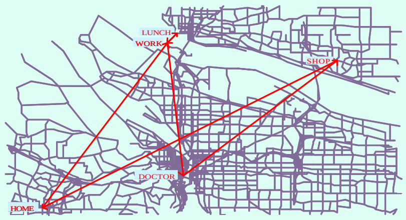


These data come from regional surveys which in general include little or no location information, at most perhaps a variable indicating the distance to the nearest light-rail stop. They do include a number of demographic variables, of which the most important for determining activity patterns are, unsurprisingly, things like the number of adults and the number of workers in the household and the ages of household members.

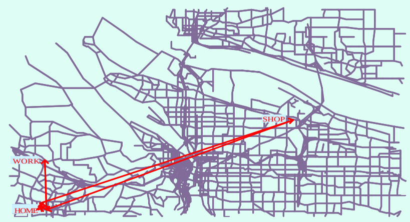
In this algorithm a regression tree is constructed for the survey data and an activity pattern, including transportation mode preferences, for a given synthetic household is chosen from the corresponding node of this tree. Note that there is no guarantee here that the transportation mode preferences assigned by this method will be efficient or even feasible for a given household. For example, individuals for whom transit is a very inefficient option, because they have no way to get to an appropriate transit terminal from their home, may nevertheless by the luck of the draw get matched to a survey household for which transit is the preferred mode of travel to work. This is the sort of inconsistency that must be remedied in the model calibration step.

Example Activities in Portland, Oregon

first person in household



second person in household



Given activity patterns, work locations are assigned with some regard to home and mode preference, using a gravity model. Other locations (shopping, school) are assigned contingent on home and work locations.

Times are merely jittered from survey times and may also require adjustment during calibration.

III. Modeling with TRANSIMS: Model Calibration

- Representation of dynamics
- Boundary conditions
- Calibration
  - Parameters in microsimulator
  - Initial conditions
  - Boundary conditions
- Assessment

Statistical simulation of boundary conditions and material properties in a model is probably one of the better developed areas of interaction between modelers and statisticians. For example, we know a lot about the geostatistical simulation of heterogeneous geological properties conditional on available observations for modeling oil reservoirs. You might think that we (as statisticians) also know quite a bit about calibration as well, but this is an area in which there is in fact much work yet to be done.

The parameters in the microsimulator, of which as you recall there were on the order of a dozen, are calibrated on small networks consisting of a few links or intersections of various types. These provide locally satisfactory behavior of simulated traffic across the range of traffic densities to be simulated, reproducing well-known density-flow relationships.

Of greater interest is "calibration", in the broad sense suggested earlier, of the coupled system in order to obtain initial and boundary conditions that are consistent with the level of service provided by the network (with the microsimulation parameters now set.)

"Spin-up" of an ocean model

Necessary to obtain an internally consistent starting point for solving a coupled system of equations forward in time:

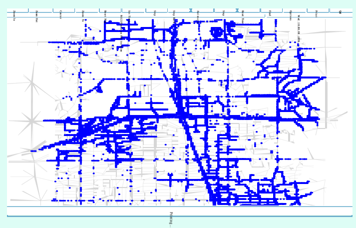
- Velocity = f (Pressure)
- Pressure = g (Salinity, Temperature)
- (Salinity, Temperature) = h(Velocity)

Returning to the analogy of the ocean model for a moment, the problem here is somewhat analogous to the spin-up of a ocean model. For an ocean model, spin-up is necessary because complete, consistent instantaneous fields of velocity, pressure, salinity and temperature are not known. From any reasonable initial pressure field, a realistic velocity field will emerge quickly by running the model, as it is an immediate consequence of the dynamics built into the equations. However, salinity and temperature, initially specified only approximately, emerge much more slowly. An ocean model must be run for hundreds or even thousands of simulated years with steady state (seasonal) forcing conditions before it relaxes into a steady, self-consistent state.

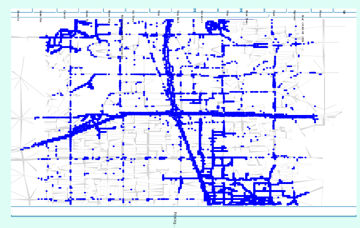
In the case of a TRANSIMS model, the "spin-up" process has to be designed much more explicitly, and it involves feedback among the modules of the model. But again, the problem is the initial lack of complete, consistent information about demand, and the goal is to determine initial (and boundary) conditions that are consistent with the dynamics and the level of service provided by the network.

Iteration in TRANSIMS

Iteration 0

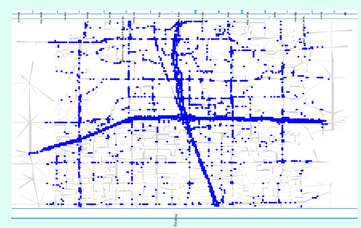


Iteration 1



feedback is required to stabilize a nonlinear system
the iteration process lets activities, route plans, and traffic converge to quasi-equilibrium

Iteration 10



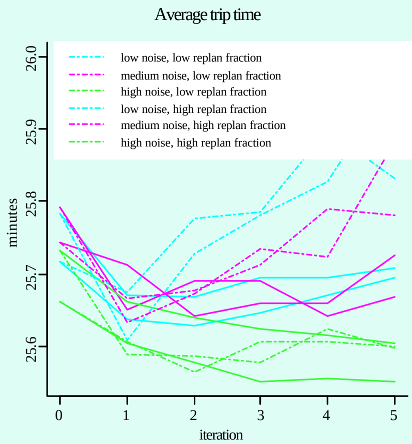
There are multiple time scales in TRANSIMS as in the ocean model. Initially the route planner, unless given other information (which might be made available from macrosimulation), will route all the travelers assuming rated performance for each link, that is, using posted speed limits. This is equivalent to assuming that there will be no other travelers using the network, that is, to ignoring dependencies among travelers. Rather than trying to parameterize such dependencies (which would be an impossible task), we allow the dynamics of the model to inform the system, exactly as the ocean modelers allow the dynamics of their model to do the work of initializing it.

Specifically, feedback from the microsimulator enables the router to do a better job of taking the time-dependent load on the network into account when planning routes for individual travelers. Realistic simulations for fixed demand scenarios can emerge relatively quickly from this process, since travel times and thus optimal routes are a fairly immediate consequence of the dynamics modeled by the Traffic Microsimulator.

Slower to emerge are adjusted patterns of demand in response to the time-dependent levels of service offered by the network.

The goal of the TRANSIMS feedback team is to come up with some generally applicable recommendations for designing efficient feedback algorithms, bearing in mind that the most time-consuming part of simulation on a large network is generally the traffic microsimulation itself.

Accelerating Feedback Convergence (Microsimulator to Router)



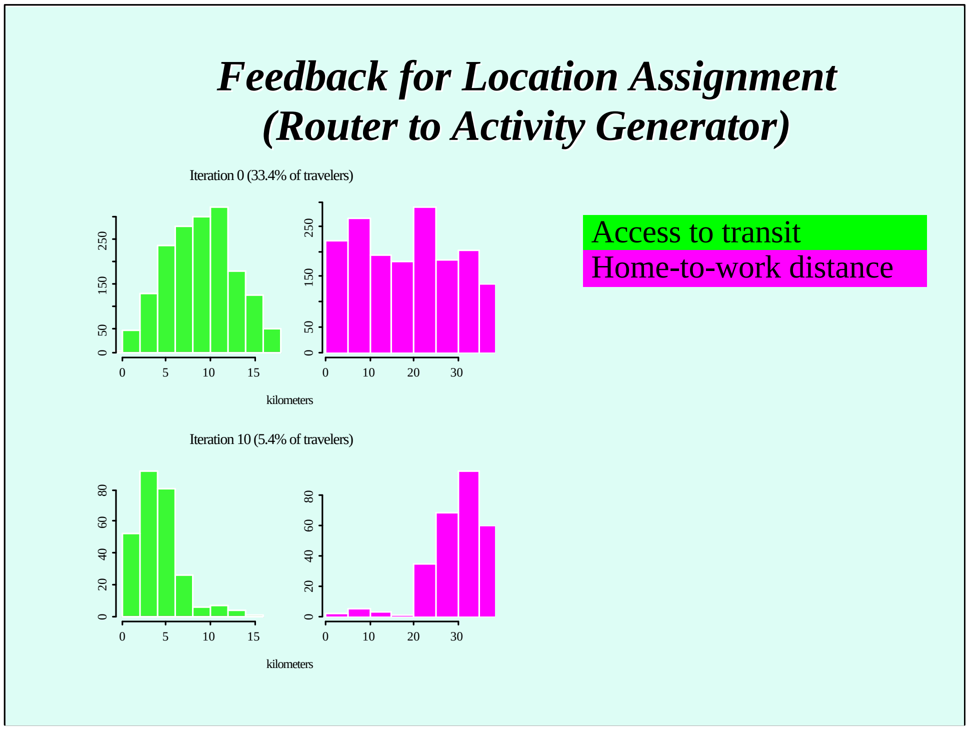
	Iteration			
	0	1	2	3
Replan fraction		0.22	0.17	0.02
Noise level	0.14	0.18	0.00	0.00
Replan × Noise		0.87	0.04	0.09

Significance level:	
	<0.05
	<0.10
	<0.20

This slide illustrates a typical result of our microsimulation-router feedback studies. The graph on the left shows the average travel time for 900 travelers on a small network. We made twelve runs with two levels of one variable, the fraction of travelers allowed to replan, and three levels of another, the amount of noise seen by the router when it looks at the average time delays on the network recorded by the microsimulator. (Routing uses a deterministic algorithm, but it can be made to act stochastically by providing it with a noisy version of network performance as estimated by the Traffic Microsimulator.) Each of the six parameter combinations was run using two different sets of random seeds, providing an error term for formal analysis of variance.

In general we find that router noise is an extremely important variable. Since the basic problem is the unrealistic expectations of the router, the most important factor in terms of accelerating convergence of the model is to ensure that on the first pass the travelers spread out across the network so that the microsimulator can provide the router with more realistic information about the effects of other travelers on the network. Notice that what we're interested in is have the model learn fast; we are not really trying to model how individuals learn to "game" the system although there may be a superficial similarity.

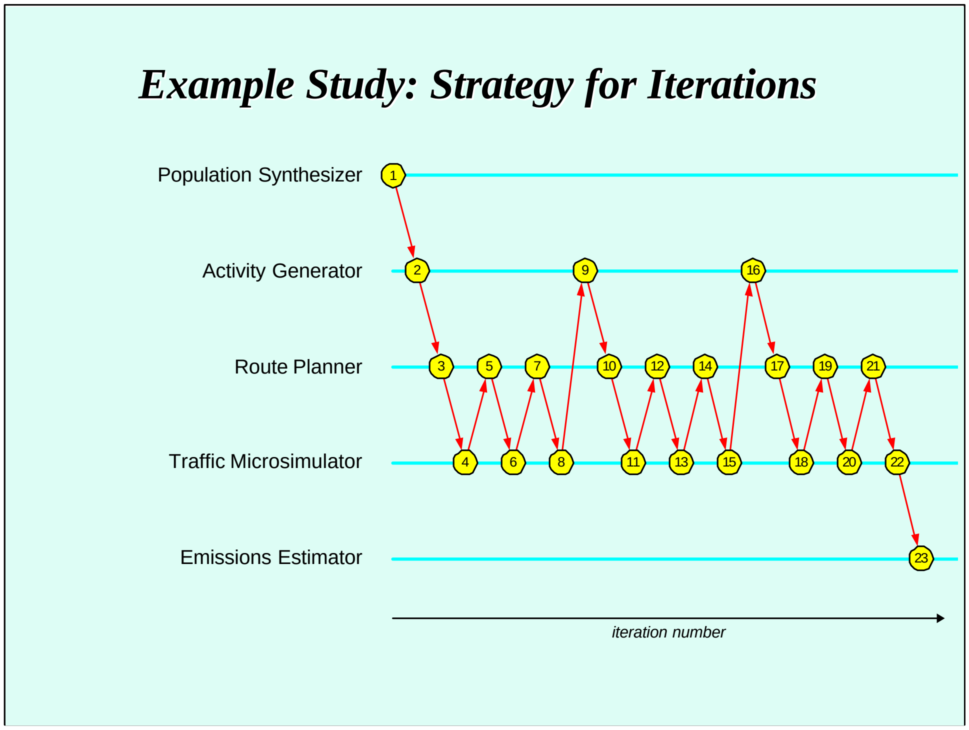
The more travelers are allowed to replan, the more important router noise becomes in later iterations as well. Results can actually diverge if too many people are allowed to replan using fairly accurate information.



We are just beginning our studies in how to use feedback from the router and the microsimulator to adjust demand patterns. Ultimately, this is a more central problem if TRANSIMS is to be used to model how travelers will change their preferences in response to changes in levels of service (new highways, new mass transit system, changes in downtown parking and pedestrian policies, etc.).

This slide is based on another simple network, where assignment of mode preference is initially random (an admittedly silly algorithm.) After several passes through the router, which identifies travelers with inefficient or impossible mode preferences (not only those whose access to transit is poor but also those whose access to transit is good because they live and work near the SAME transit stop), we arrive at a much more reasonable distribution of travelers who use transit (and who walk) on this network.

Of course in the real case we can use much more sophisticated algorithms to make the initial assignments and subsequent reassignments of travelers. However, one of the main goals of TRANSIMS is to be able to model how travelers will change their preferences in response to changes in levels of service, so it is important to have this feedback loop in the model to let the model dynamics refine this boundary condition, no matter how sophisticated such assignment algorithms may be.



The results of our studies so far suggest that good feedback strategies will probably spend more time between the route planner and the activity generator, with much less frequent invocation of the traffic microsimulator than suggested by this older slide.

This part of my talk has illustrated a second area in which statisticians and modelers can interact, particularly in constructing coupled models. There is a role in optimizing the coupling, as I have illustrated. This entails the application of standard and nonstandard statistical tools for the design and analysis of computer experiments.

IV. Modeling with TRANSIMS: Model Assessment

- Representation of dynamics
- Boundary conditions
- Calibration
- Assessment
  - Selection and (usually) transformation of available data
  - Generation and (usually) manipulation of computer output
  - Statistical inference

Finally we come to the ultimate challenge. Models can be constructed for several purposes, and in the end we need to assess the model against some criteria that will depend on the proposed use of the model.

Model assessment questions

- Does the model capture our understanding of the process being modeled?
  - Does it reproduce some important observable features of this process?
- Does it augment that understanding or provide us with new information?
  - E.g., what drives the model? the process? And are they the same?
- Can we use it to extrapolate beyond observable conditions?
  - How much confidence should we attach to model predictions?

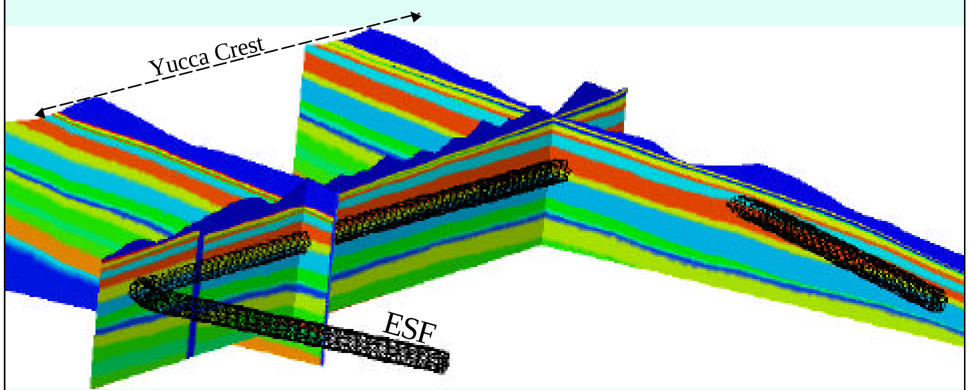
Here are some of the questions we may be asking at this stage. Along the lines of the first of these, I've already alluded to a model's ability to capture emergent ensemble dynamics which are not built into the model explicitly. In an ocean model, accurate simulation of the Gulf Stream is a common goal when modeling the North Atlantic. In TRANSIMS, we are able to model how traffic changes under congested conditions, which is where traffic macrosimulation models fail.

Along the lines of the second question, we are beginning to learn some things from our feedback studies. For example, perfect information is almost certainly not desirable in terms of optimizing the system, something that should be taken into account when designing Intelligent Transportation Systems.

The big question is, how do we go about "model accreditation", that is, building confidence in its usability for the sorts of questions that, say, Municipal Planning Organizations need to answer?

The answer is probably: incrementally. Good tests of the model should look for data and phenomena that were not used in the construction or calibration of the model, and we should not expect to be able to devise a test or even a series of tests that can address all our uncertainties.

Yucca Mountain Model



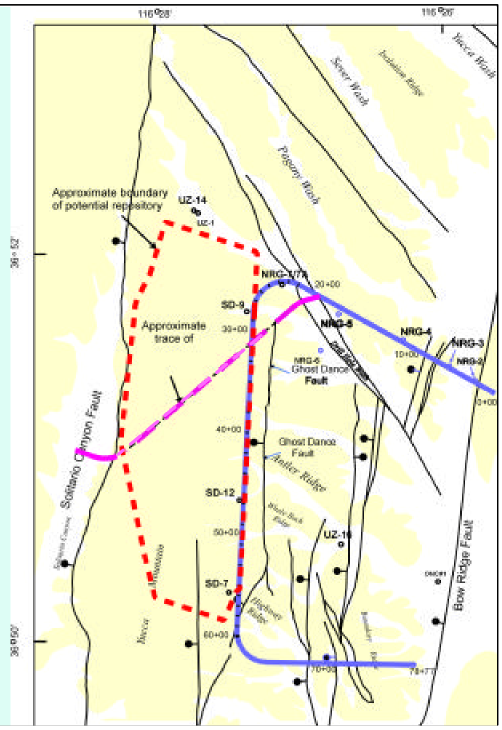
Goal: Performance assessment for a potential geologic repository for high-level nuclear wastes

Since we haven't done too much formal assessment work on TRANSIMS models yet, I propose to digress briefly to a similar problem in another context, performance assessment at Yucca Mountain. Here the problem is to predict the likelihood of success of a geologic repository for 10 thousand, 20 thousand or more years into the future, in sequestering the wastes stored there from the accessible environment.

For this purpose, a large and fairly detailed performance assessment model of flow and transport at Yucca Mountain has been created, and much work has gone into refining the parameterization of this model. Most of the extensive data collected has been assimilated into this model.

Yucca Mountain Data:

About 200 chlorine-36 samples from the Exploratory Studies Facility, a tunnel (solid blue line) collected at the depth of the potential nuclear waste repository (dashed red line)

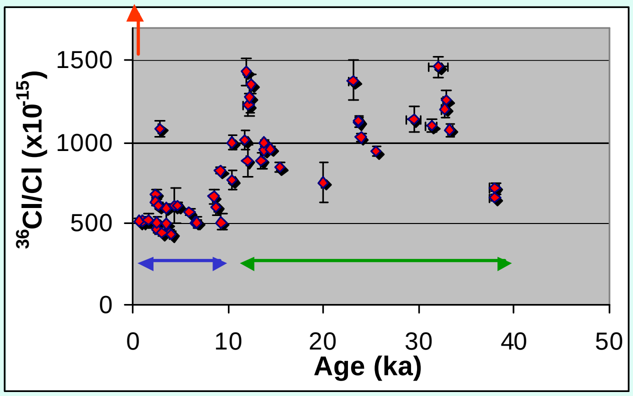


An exception are some data on chlorine-36, which is both a cosmogenic and an anthropogenic isotope in the atmosphere. It is a radioactive isotope with a half-life of about 300 thousand years. Chlorine is removed rapidly from the atmosphere by precipitation and travels underground with the water in which it is dissolved without sorbing onto the substrate, so it is potentially a good natural tracer for subsurface flow processes.

Historic ^{36}Cl/Cl Source Ratio from Plummer et al. (1997)

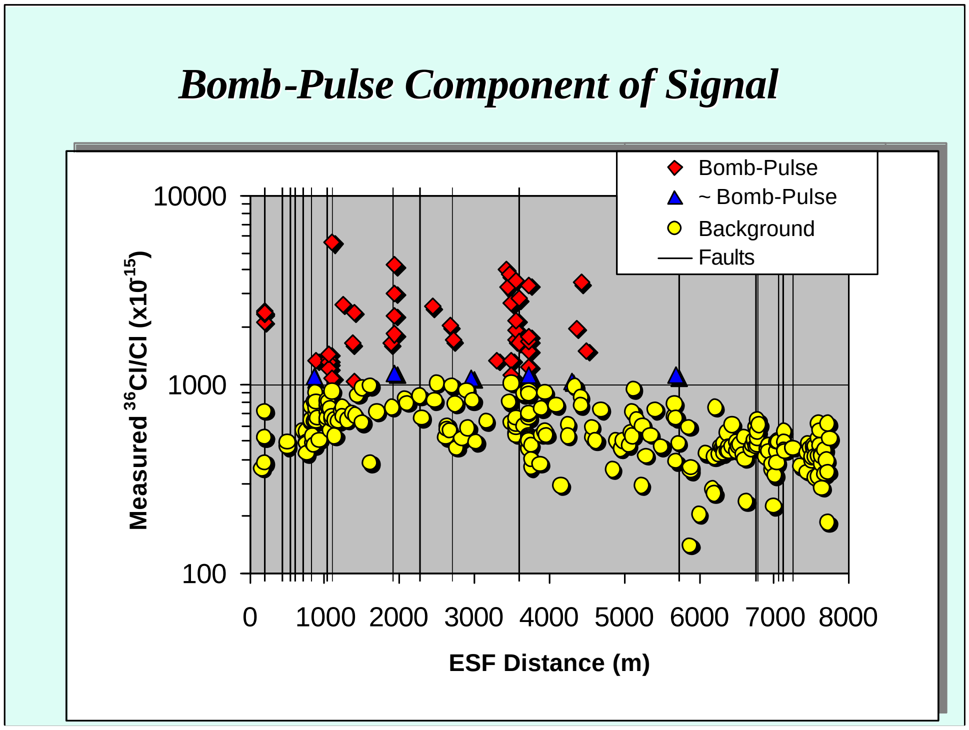
<u>Three Primary Components:</u>

- Bomb-pulse less than 50 years ago.
- Fairly constant Holocene signal.
- Elevated signal at end of Pleistocene.



What makes chlorine-36 valuable in this context is the fact that its concentration has not been constant either historically or in geologic time. Note that what is actually measured is the ratio of chlorine-36 to total chlorine; the radioactive isotope constitutes only a minute fraction of the chlorine in the atmosphere. There was a fairly sharp shift downward in this ratio, by a factor of about two, at the end of the Pleistocene, about 10 thousand years ago. There was also a brief period of extreme elevation, more than two of orders of magnitude above the Holocene background, caused by atmospheric testing in the late 1950s.

So what we observe at depth at Yucca Mountain is the outcome of a natural experiment that has been going on over a length of time comparable to the time spans of interest for performance assessment. This is quite different from what we can get from designed experiments at either the laboratory or even the field scale, although those which are underway now will give us different information than we can extract from these data.



This is what is observed at the level of the repository. (Some of the data at each end are from locations closer to the surface.) From the last slide we know that ratios above 1500 times 10-15 are above the background range of even the Pleistocene, so it is clear that we are seeing some very recent water in some of these samples. To some extent we can associate these spikes with major faults that cut through an overlying stratum in which flow is otherwise dominated by slow matrix flow. Excluding these obvious samples, however, there is still apparently a trend in the data along the tunnel.

Two explanations for trend

- Above the north end of the tunnel, the PTn retards flow sufficiently that substantial fraction of water is of Pleistocene age
  - Constrains infiltration rates to low end of possible
- Most of the samples in the north end are either below through-cutting faults or in joints that could be connected to such faults by low-angle joints
  - Suggests need for better parameterization of fracture-dominated flow in welded tuff strata

There are two possible explanations for this trend. Over the north end of the tunnel, the nonwelded PTn layer is much thicker than at the south end, which certainly reduces the average flow rate from the surface. Whether it reduces it sufficiently to lead to a substantial component of Pleistocene-age water in samples collected in the ESF is less clear. But the preceding slide also indicates the association of the highest, indisputably "bombpulse" samples with major faults through the PTn, and contingency table analysis suggests that where "bombpulse" is found away from such faults it may well be primarily in joints that are connected to such faults by low-angle joints. In part because sampling in the tunnel has been highly biased toward geologically "interesting" features, there are few samples that are clearly away from both faults and joints.

Andy Wolfsberg and I have been exploring the data in the context of the model and the model in the context of the data to try to determi ne which is the predominant effect. To the extent that our results may end up influencing improvements in the design of the performance assessment model, they might be considered part of the calibration process. But that the model, with all of its current limitations and uncertainties, can produce results that are reasonably consistent with these data should support its use as one source of information for decision making concerning the potential repository. Of course this work does not address many critical aspects of performance, notably transport. Like most "validation" tests, it is thus far fromcomplete.

Models and Statistics: Two views into the problem

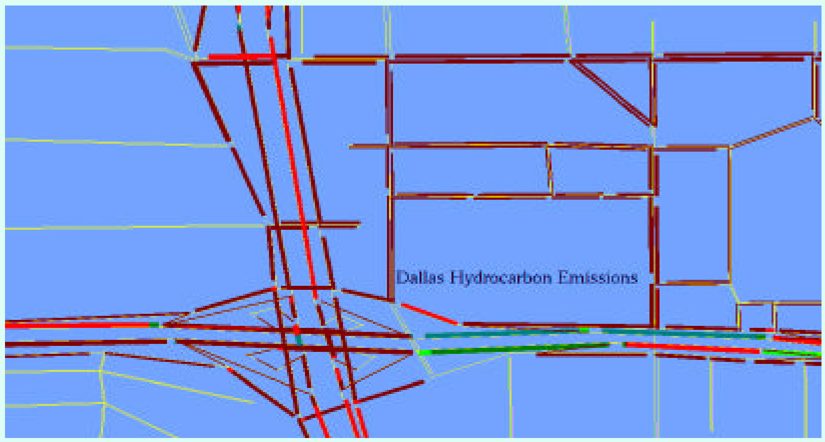
- Model limitations
  - Fractures handled by dual permeability/dual porosity type modeling
  - Heterogeneity within strata is not modeled
- Data limitations
  - Observations are noisy mixtures of several components
  - Geology only partially observed (low-angle intersecting joints above tunnel are unobservable)
- Model + Data synergy?

The jury is still out on these questions, but my points are two:

The first is that the model has known and potentially serious limitations. The data have even more obvious limitations. But by putting them together we can learn more than we can learn from either separately.

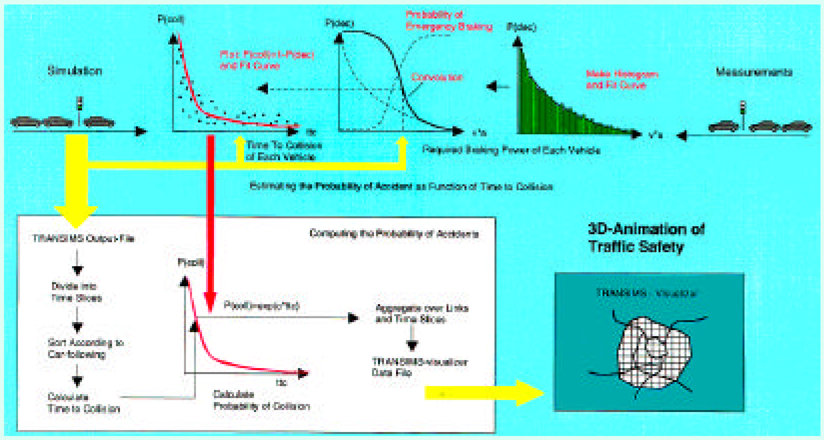
And the second is that we need to look carefully at the possible tests we can devise for our model, not confusing model "validation" (I prefer the word "assessment", because "validation" as commonly understood is probably beyond our abilities) with the model calibration step.

Example Hydrocarbon Emissions in Dallas, Texas



Where do we look for such tests of TRANSIMS models? One place is in the final module of the TRANSIMS system, which looks at detailed acceleration patterns in the microsimulation and extracts estimates of emissions which are localized in both space and time. These can be compared (undoubtedly only after further, statistical post-processing) with monitoring information.

Simulation of Probability of Accidents



A second possibility that is being worked on is using the output of the microsimulator to estimate space-time patterns of accidents, something else on which there is typically quite a bit of data. As for emissions, this requires quite a bit of post-processing of the microsimulation data, where statisticians might again play a role.

Both of these are regional-scale tests, which is what we need if we are proposing the model for regional-scale use. And both use aspects of the model output that were <u>not</u> explicitly parameterized in building the model; that is, patterns of emissions and accidents are again “emergent properties” of the model. To the extent that they correspond with observations on the same phenomena, a potential user may feel more confident in making use of other, less verifiable, model predictions for planning purposes, although of course such tests do not “validate” the model in any commonly understood sense of that word.

Yet another possibility, exercising a different aspect of the model, might be to simulate travelers’s reactions to a planned closure of part of the network for construction over a period of several months (long enough for the daily travelers to adjust their demand patterns.)

V. Conclusions

- Despite the novelty of the TRANSIMS architecture, modeling with TRANSIMS requires the same types of decisions and analyses as modeling with more "traditional" systems.
- Major statistical development is needed, for example,
  - Using models in "inverse" mode to refine (calibrate) input fields (e.g., Kennedy and O'Hagen, Glimm et al., Raftery, …)
  - Standardize to extent possible recommendations and tools for building confidence in models, particularly if they are to be used in decision making or policy contexts.

So I hope I have accomplished two things in this talk:

First, I hope that I have convinced you that despite the novelty of the construction of TRANSIMS, models built with this system will have pretty much the same set of statistical problems that can be addressed with pretty much the same set of tools as more traditional types of computer models.

And second, I hope I have convinced you that while some of these tools are in place, many many others need to be developed or refined in order to address these problems. This is why I think that statistical work with models and modelers is really at the heart of the Interface between Computing and Statistics.